

## Computing invariant measures for expanding circle maps

Michael Keane<sup>†||</sup>, Rua Murray<sup>‡¶</sup> and Lai-Sang Young<sup>§+</sup>

<sup>†</sup> Centrum voor Wiskunde en Informatica, PO Box 94079, 1090 GB, Amsterdam, the Netherlands

<sup>‡</sup> Statistical Laboratory, University of Cambridge, 16 Mill Lane, Cambridge CB2 1SB, UK

<sup>§</sup> Department of Mathematics, University of California, Los Angeles, CA 90095, USA

Received 6 February 1997

Recommended by P Grassberger

**Abstract.** Let  $f$  be a sufficiently expanding  $C^2$  circle map. We prove that a certain *Markov approximation scheme* based on a partition of  $S^1$  into  $2^N$  equal intervals produces a probability measure whose total variation norm distance from the exact absolutely continuous invariant measure is bounded by  $CN2^{-N}$ ;  $C$  is a constant depending only on the map  $f$ .

AMS classification scheme numbers: 58F11, 28D99, 65U05

### 1. Introduction

For 20 years it has been known that certain Markov approximations to dynamical systems are useful for calculating invariant measures [5]. In this paper, we give quantitative bounds on the rate of this approximation. While the abstract results in [5, 2, 3] and many others focus on the convergence of the scheme in the limit, our approach is to actually bound the accuracy of the approximation for a fixed approximant. The result that we prove (theorem 1) is stated in section 1.3 and provides precisely such a bound; the convergence of the scheme to the invariant measure follows immediately, as explained in section 1.4.

Because we focus on a fixed approximant, our method of proof entails a careful analysis of the mixing properties of the particular Markov chain (theorem 2). This part of the proof is in section 3, and section 2 consists of preparatory observations for this proof; both sections may be omitted at a first reading. Section 4 connects the approximating Markov chain with the exact invariant measure, and contains the remainder of the proof of theorem 1. A concrete numerical application is provided in section 5.

#### 1.1. Computing invariant measures

To apply ergodic theory to a dynamical system, it is necessary to have an invariant measure. Occasionally, an appropriate invariant measure can be obtained analytically; generally one will require a numerical procedure to get a 'good' approximation. Because the space of probability measures on a general phase space is infinite dimensional, this is typically a difficult problem. Often, one is interested in *physical* or SBR-measures; these invariant measures describe the asymptotic statistical behaviour of a 'large' set of initial points.

<sup>||</sup> E-mail address: M.S.Keane@cwi.nl

<sup>¶</sup> E-mail address: R.Murray@statslab.cam.ac.uk

<sup>+</sup> E-mail address: lsy@math.ucla.edu

Consequently, the usual technique for approximating an invariant measure is to simulate a long orbit, and weight regions of phase space according to how often the trajectory visits. In favour of this method, it is easy to show that if the orbit is exact, then the approximation converges weakly to an invariant measure as the orbit length tends to infinity. Against it, there are the facts that it is rarely possible to simulate an orbit exactly, that the underlying theory gives no information about the *rate* at which the limiting measure is approached, and that there is no *a priori* reason to suppose that the asymptotic measure is of physical interest.

We want a technique which does not have these drawbacks. That is, we seek to simultaneously control which invariant measure we are converging to, while having concrete estimates on the rate of approximation. The Markov approximation method described below has these properties.

### 1.2. Area overlaps Markov chain

A relatively well known approximation method is called the *area overlaps Markov chain*. The idea is to approximate the map globally by a Markov chain, and use the invariant probability measure for the Markov chain as an estimate of the physical invariant measure of the dynamical system. We will work with the circle  $S^1$ , equipped with the Borel  $\sigma$ -algebra and normalized Lebesgue measure  $m$ . Let  $f$  be a mapping of  $S^1$  to itself. Our interest is in the case when the dynamical system  $(S^1, f)$  has an *absolutely continuous invariant probability measure*  $\mu$  (ACIM). If  $\mu$  is an ACIM, then

$$\mu(f^{-1}A) = \mu(A) \quad \text{and} \quad \mu(A) > 0 \Rightarrow m(A) > 0$$

for every Borel set  $A$ .

Next, we describe the approximation scheme. Fix  $N > 0$ . Divide the circle into  $2^N$  equal subintervals:  $\{X_i\}_{i=1}^{2^N}$ . Then for each pair  $i, j$ , put

$$P_{ij} = \frac{m(f^{-1}X_j \cap X_i)}{m(X_i)}.$$

Because  $m$  is a probability measure, and  $\{X_i\}_{i=1}^{2^N}$  is a partition of  $S^1$ , we have that  $\sum_{j=1}^{2^N} P_{ij} = 1$  for each  $i$ . Hence, the  $(P_{ij})$  define a family of *transition probabilities* on the partition elements. The approximation procedure is to solve for an invariant probability measure for the Markov chain governed by  $P$ . We denote this measure by  $p = p^{(N)}$ , and have the approximation:  $p_i \approx \mu(X_i)$ . (We have called  $P$  the *area overlaps Markov chain* because  $P_{ij}$  is precisely the proportion of Lebesgue mass ('area') in  $X_i$  which is mapped into  $X_j$  under one iteration of  $f$ .)

This Markov approximation scheme has a long history: Originally suggested by Ulam [10], it has appeared in various forms in the literature. Li [5] proved that as the partition is refined, the approximate measures converge strongly to the acim for the case of a one-dimensional expanding map. Ding and Zhou [2] obtained a partial extension of this result to multidimensional expanding maps, and Froyland [3] proved a similar result for weak convergence to the SBR-measure of Anosov systems. However, none of these authors obtained estimates of the *rate* of approximation. Our work fills this gap because theorem 1 provides a *quantitative bound* on the distance of  $p^{(N)}$  from the exact invariant measure  $\mu$ .

### 1.3. The theorem

Let  $f : S^1 \rightarrow S^1$  be a uniformly expanding circle map. We give precise conditions by working with a *lift* of  $f$ : Despite the possible ambiguity, we also denote the lift to  $\mathbb{R}$  by  $f$ . Then,

$$f(x) = f(x + n) \bmod 1$$

for all  $x \in [0, 1)$ ,  $n \in \mathbb{Z}$  (for convenience, we assume that the  $\mathcal{C}^2$  map  $f$  is *increasing* as a function of  $\mathbb{R}$ ). We require that there exists a constant  $\lambda > 4$  such that for each  $x \in [0, 1)$

$$|f'(x)| \geq \lambda. \quad (1)$$

This is the *expansivity* condition. (Note that any  $\mathcal{C}^2$  circle map with derivative strictly bounded away from 1 can be turned into an  $f$  of our form by taking sufficiently many iterates.) Usually,  $\lambda = \min_{x \in S^1} |f'(x)|$ , and we denote for use below  $\lambda_{\max} = \max_{x \in S^1} |f'(x)|$ . Finally, we impose a type of *bounded distortion* condition by requiring that  $\log |f'|$  is Lipschitz continuous. That is, there exists a constant  $K > 0$  such that for any  $x, y \in [0, 1)$

$$|\log |f'(x)| - \log |f'(y)|| \leq K\rho(x, y) \quad (2)$$

(the metric  $\rho$  on  $[0, 1)$  is defined as usual by  $\rho(x, y) = \min\{|x - y|, 1 - |x - y|\}$ ). Any  $\mathcal{C}^2$  map satisfying (1) automatically satisfies (2) for some constant  $K$ .

Because we are interested in computing ACIMs, we need the following.

**Basic fact (folklore theorem).** *Let  $f : S^1 \circlearrowleft$  satisfy (1) and (2). Then there exists a unique absolutely continuous invariant measure  $\mu$ ; moreover, the corresponding density  $\phi$  is such that  $\log \phi$  is Lipschitz continuous. In particular, if  $x, y \in S^1$  then*

$$\left| \log \frac{\phi(x)}{\phi(y)} \right| \leq \frac{K}{\lambda - 1} \rho(x, y).$$

**Remark 1.1.** The folklore theorem dates back to Renyi [9], but the smoothness of the density that we use is more evident from the proofs in [4] or [11].

**Theorem 1 (the main result).** *Let  $f$  satisfy (1) and (2), let  $\mu$  be the acim which is guaranteed to exist by the folklore theorem. Suppose that the partition  $\{X_i\}$  of  $S^1$  consists of  $2^N$  equal subintervals,  $P$  is the corresponding area overlaps matrix,  $p$  is an invariant probability vector for  $P$  and that the vector  $q$  has components  $\mu(X_i)$ . Then there exists a natural number  $N_0$  (a formula for  $N_0$  is provided below) and constant  $C > 0$  such that whenever  $N \geq N_0$*

$$\|q - p\|_1 \leq C \frac{N}{2^N}.$$

The constant  $C$  has an explicit formula in terms of  $N_0$ ,  $K$  and  $\lambda$  and is thus independent of  $N$ .

**Remark 1.2.** The theorem is proved as theorem 3 in sections 3 and 4. We use

$$N_0 = \lceil \max\{\log \lambda_{\max} / \log 2, \log 3 / \log(\lambda/4)\} \rceil.$$

The constant  $C$  is explained in remark 4.1 (following theorem 3), and depends on the choice of  $N_0$ . In remark 4.1, we show that

$$C = C(N_0) \leq \frac{K}{\lambda - 1} e^{K/(\lambda-1)} + \epsilon(N_0)$$

where the function  $\epsilon(N_0) \rightarrow 0$  exponentially fast as  $N_0 \rightarrow \infty$ . Therefore, using a larger value of  $N_0$  than specified above, one can obtain a much better bound on the approximation error.

We give some idea of this with an example. Suppose that  $f$  is a  $C^2$  circle map with  $\lambda = 5$ , and  $|f''| \leq 20$ . It follows easily that  $K = \frac{20}{5} = 4$  will satisfy (2), and that  $\lambda_{\max} \leq 25$ . Consequently, the minimal choice of  $N_0$  is 5. Because 5 corresponds to a very coarse partition of  $S^1$  (each set has length  $\frac{1}{32}$ ), we illustrate the result for  $N_0 = 8, 10$  (these still correspond to rather coarse partitions, for which the computation of an invariant vector  $p$  takes a few seconds). From remark 4.1, we obtain values  $C(8) = 28.7$ ,  $C(10) = 6.65$ . (In this case  $C(N_0) \rightarrow 2.71$  as  $N_0 \rightarrow \infty$ .)

#### 1.4. Strong convergence to acim

Strong convergence of the Markov approximation scheme follows from theorem 1. Let  $p^{(n)}$  be the invariant probability vector for the Markov approximation on a uniform partition  $\{X_i\}_{i=1}^{2^n}$ , and let  $\psi^{(n)} \in L^1(S^1)$  be defined by

$$\psi^{(n)} = \sum_{i=1}^{2^n} 2^n p_i^{(n)} 1_{X_i}.$$

$\psi^{(n)}$  is called the *density corresponding to  $p^{(n)}$* . The metric  $d$  on  $L^1(S^1)$  defined by

$$d(g_1, g_2) = \sup_{\{h \in L^1: |h| \leq 1\}} \int_{S^1} (g_1 - g_2) h \, dm$$

induces the topology of *strong convergence*. We prove strong convergence of the Markov approximation scheme by showing that  $d(\psi^{(n)}, \phi) \rightarrow 0$  as  $n \rightarrow \infty$  (where  $\phi$  is the invariant density from the folklore theorem).

Let  $q^{(n)}$  be the probability vector with components  $q_i^{(n)} = \mu(X_i)$ , and let  $\phi^{(n)}$  be the density corresponding to  $q^{(n)}$ .

It is easy to check that

$$d(\psi^{(n)}, \phi^{(n)}) = \int_{S^1} |\psi^{(n)} - \phi^{(n)}| \, dm = \|p^{(n)} - q^{(n)}\|_1$$

(where  $\|\cdot\|_1$  is the usual vector 1-norm on  $\mathbb{R}^{2^n}$ ). Moreover, it follows from the folklore theorem that

$$d(\phi^{(n)}, \phi) \leq e^{K/(\lambda-1)} (e^{2^{-n}K/(\lambda-1)} - 1).$$

Hence,

$$d(\psi^{(n)}, \phi) \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

## 2. Notation

Before proceeding to the proof of the theorem, we need to define what is meant by the *lift* of the Markov chain governed by  $P$ , and establish some notation.

### 2.1. Lift of the Markov chain

We construct the *lift* of the Markov chain from the lift of  $f$ .

Recall that the stochastic matrix  $P$  was obtained by putting

$$P_{ij} = \frac{m(f^{-1}X_j \cap X_i)}{m(X_i)}$$

where  $f$  is regarded as a map of the circle, and the sets  $X_i = [\frac{i-1}{2^N}, \frac{i}{2^N})$  partition  $S^1$ . This induces a Markov chain on the integers  $\mathbb{Z}_{2^N} = \{1 \dots 2^N\}$ . Totally analogously, we partition  $\mathbb{R}$  into intervals of length  $2^{-N}$ , indexed by the integers  $\mathbb{Z}$ . Using this partition, and the lifted map  $f$  (as a function of  $\mathbb{R}$ ), we can define a Markov chain on the integers  $\mathbb{Z}$  according to the same formula, with the understanding that  $m$  then denotes the Lebesgue measure on  $\mathbb{R}$ . This will be the *lift* of the Markov chain governed by  $P$ .

Throughout the paper, elements of  $\mathbb{Z}$  and  $\mathbb{Z}_{2^N}$  will be denoted by lower case letters. If there is possible ambiguity, then elements of  $\mathbb{Z}_{2^N}$  will be denoted with a  $\hat{\cdot}$ . Then,  $k \in \mathbb{Z}$  will be written as  $\hat{k} + l2^N$  for some integer  $l$ ,  $\hat{k} \in \mathbb{Z}_{2^N}$ .

We need a notation to describe the lifted Markov chain: Let the  $n$ -step transition probabilities in the lift be denoted by

$$P_n(j; i) = Pr(x_n = j | x_0 = i) \quad (3)$$

where  $x_0, x_1, \dots$  is a sequence of the lifted Markov chain on  $\mathbb{Z}$ .

To ensure that the notation for the lift is in unambiguous agreement with the finite-state Markov chain, we insist that the partition over which  $P$  is constructed is sufficiently fine that no subinterval  $X_i$  can map over the entire circle  $S^1$  under one iteration of  $f$ . By the mean-value theorem, this is guaranteed if  $|f'(x)|2^{-N} < 1$  for each point  $x \in S^1$ . Since  $N_0 \geq \log_2 \lambda_{\max}$  (cf remark 1.2), this condition is satisfied whenever  $N \geq N_0$ . Therefore, if  $P_1(j; i) > 0$ ,

$$P_1(j; i) = P_{\hat{j}\hat{i}}$$

where  $\hat{i} \equiv i \pmod{2^N}$ , and  $\hat{j} \equiv j \pmod{2^N}$ . With this in mind, we will often indulge a slight abuse of notation by writing  $P_1(j; i) = P_{ij}$ .

Our strategy in section 3 is to derive uniform mixing estimates for the matrix  $P^N$  by obtaining uniform (in  $i, j, n$ ) quantitative control of  $P_n(j; i)$ . Consequently, we also define a notation for the support of each probability distribution  $P_n(\cdot; i_0)$ :

$$\mathcal{R}_n(i_0) = \{j \in \mathbb{Z} : P_n(j; i_0) > 0\}. \quad (4)$$

In other words,  $\mathcal{R}_n(i_0)$  consists precisely of those states in the lift which are *reachable* in  $n$  steps from the state  $i_0$ . In trying to get uniform estimates on the distribution  $P_n(\cdot; i_0)$ , it will be useful to consider those states which are reachable in  $n$  steps from starting at  $i_0$ , and not reachable in  $n$  steps from any other state. We denote this subset of  $\mathcal{R}_n(i_0)$  by

$$\mathcal{R}_n^*(i_0) = \mathcal{R}_n(i_0) \setminus \{\cup_{i \neq i_0} \mathcal{R}_n(i)\}. \quad (5)$$

For example,  $\{j\} \in \mathcal{R}_1^*(i_0)$  if and only if  $X_j \subset f(X_{i_0})$ , whereas  $\{j\} \in \mathcal{R}_1(i_0)$  if and only if  $m(X_j \cap f(X_{i_0})) > 0$ .

Finally, we make the obvious definition

$$P_n(\Lambda; i) = \sum_{k \in \Lambda \cap \mathcal{R}_n(i)} P_n(k; i)$$

for a subset  $\Lambda \subset \mathbb{Z}$ . Since our ultimate interest is in projecting distributions on the lift back down to distributions on  $\mathbb{Z}_{2^N}$ , we observe the important identity

$$(P^n)_{ij} = \sum_{\{k=j+l2^N : l \in \mathbb{Z}\}} P_n(k; i) \quad (6)$$

for  $i, j \in \mathbb{Z}_{2^N}$ .

### 3. Quantitative mixing properties of the Markov approximation

We show below that the expansivity and regularity properties of the map  $f$  induce similar properties in the area overlaps matrix  $P$ . This allows us to prove a *uniform distortion property* for rows of the matrix  $P^N$  (recall that  $P$  is a  $2^N \times 2^N$  matrix). We then deduce an *explicit bound*, independent of  $N$ , for the speed at which  $P^N$  mixes to equilibrium (theorem 2).

#### 3.1. Regularity properties of the matrix $P$

In this section, we state some key properties of the Markov matrix  $P$ . The purpose of this is twofold: (i) to gather together all of the facts that we need for the proof of theorem 2; (ii) to show how these facts can be seen as discrete versions of the hypotheses on the map  $f$ . For an arbitrary Markov chain, these conditions will be highly unrealistic; for a Markov chain whose transition probabilities are derived from a ‘well behaved’ dynamical system by a scheme like ours, the kinds of conditions we state are more natural.

Below,  $K$  and  $\lambda$  are as for the map  $f$  (cf equations (1) and (2)). To simplify the notation, we put  $\kappa = \frac{\lambda+2}{\lambda}$ . Moreover, in (H1)–(H3) we assume that  $N \geq N_0$ , where  $N_0$  is defined in remark 1.2. This guarantees that all the statements below are well defined.

**Hypothesis H1 (continuity).** For each state  $i \in \mathbb{Z}$ , we require that

$$\mathcal{R}_1(i) = \{j : P_{ij} > 0\}$$

is an *interval* in  $\mathbb{Z}$  of length greater than or equal to  $\lambda$ . An interval in  $\mathbb{Z}$  is a sequence of consecutive integers, and the length of such an interval is the number of integers in the sequence; our hypotheses imply that each state  $i$  must have at least five possible images under the action of the Markov chain.

**Proof that  $P$  satisfies (H1).** Let  $X_i$  be the partition element corresponding to  $i$ . Then

$$\mathcal{R}_1(i) = \{j : m(X_j \cap fX_i) > 0\}.$$

Since  $f$  is continuous, and  $f'(x) \geq \lambda$  at every point of  $X_i$ , it follows that  $f(X_i)$  is an interval of length  $\geq 2^{-N}\lambda$ , and hence that  $\text{card}\{\mathcal{R}_1(i)\} > \lambda$ .  $\square$

**Hypothesis H2 (endpoint overlaps).** In comparing the sets  $\mathcal{R}_1(i)$  and  $\mathcal{R}_1(i \pm 1)$  we need to be careful. With this in mind, we require that for each  $i$ , the set

$$\mathcal{R}_1(i) \cap \mathcal{R}_1(i+1)$$

consists of *at most* one state ( $i+1$  is considered modulo  $2^N$ ).

**Proof that  $P$  satisfies (H2).** If  $X_i, X_{i'}$  are adjacent intervals in the partition, then

$$\mathcal{R}_1(i) \cap \mathcal{R}_1(i') = \{j\}$$

where  $X_j$  is the unique interval containing the image under  $f$  of the mutual endpoint of  $X_i$  and  $X_{i'}$ .  $\square$

We need the transition probabilities  $P_{ij}$  to be ‘slowly varying’ in  $i$ . It will turn out to be convenient to express this in terms of a matrix  $M$ , derived from  $P$ . The matrix  $M$  will generally not be stochastic, and is introduced entirely for the purpose of simplifying the ‘book-keeping’ in our proof.

Fix a state  $i$ . In view of (H2) there exist unique states  $j_-$  and  $j_+$  satisfying

$$\mathcal{R}_1(i-1) \cap \mathcal{R}_1(i) = \{j_-\} \quad \text{and} \quad \mathcal{R}_1(i) \cap \mathcal{R}_1(i+1) = \{j_+\}.$$

We define  $M$  by  $M_{ij_-} = P_{(i-1)j_-} + P_{ij_-}$ ,  $M_{ij_+} = P_{ij_+} + P_{(i+1)j_+}$ ,  $M_{ij} = P_{ij}$  otherwise (in the last case,  $\{j\} \in \mathcal{R}_1^*(i)$  and hence  $X_j \subset f(X_i)$ ). The idea is that we have changed the elements of  $P$  which correspond to ‘interval endpoints’ in the partition. To help with understanding, the reader should be convinced before proceeding that

$$M_{ij} = \frac{m(f_i^{-1}X_j)}{m(X_i)} \quad \text{if } P_{ij} > 0 \quad \text{and} \quad M_{ij} = 0 \text{ otherwise}$$

where the notation  $f_i^{-1}$  denotes the branch of  $f^{-1}$  which maps over  $X_i$  (recall that  $P_{ij} > 0$  if and only if  $f(X_i)$  intersects  $X_j$  in a set of positive measure). The quantity  $M_{ij}$  should be thought of as a rate of contraction for the branch of the inverse that maps  $X_j$  to  $X_i$ , rather than as a transition probability out of  $X_i$ .

**Hypothesis H3 (uniform expansion).** For every  $i, j \in \mathbb{Z}_{2^N}$ ,

$$M_{ij} \leq \frac{1}{\lambda}. \quad (7)$$

If  $M_{i_1j_1}, M_{i_2j_2} > 0$ , then

$$\frac{M_{i_1j_1}}{M_{i_2j_2}} \leq e^{\kappa \frac{|i_1 - i_2| + \kappa}{2^N}}. \quad (8)$$

Furthermore, because we are working with the *lift* of the Markov chain, we can deduce from (7) and (H1) that if  $M_{i_1j_1}, M_{i_2j_2} > 0$ , then

$$|i_1 - i_2| \leq \left\lceil \frac{|j_1 - j_2| + 1}{\lambda} \right\rceil. \quad (9)$$

**Proof that  $P$  satisfies (H3).** By expansivity, the statement in (7) follows from the claim that for each  $M_{ij} > 0$  there exists  $x \in \mathbb{R}$  such that

$$M_{ij} = \frac{1}{|f'(x)|} \quad \text{and} \quad \rho(x, X_i) \leq \frac{2^{-N}}{\lambda}.$$

**Proof of the claim.** The mean-value theorem implies that

$$P_{ij} = \frac{m(f^{-1}X_j \cap X_i)}{m(X_i)} = \frac{1}{|f'(x_{ij})|} \frac{m(X_j \cap fX_i)}{m(X_j)}$$

for some  $x_{ij} \in f^{-1}X_j \cap X_i$  (we have also used the fact that  $m(X_i) = m(X_j)$ ). Hence, the claim follows immediately with  $x = x_{ij}$  if  $M_{ij} = P_{ij}$  (because in that case  $X_j \subset f(X_i)$ ). For the other case, suppose that  $X_i$  and  $X_{i'}$  are adjacent intervals such that  $\mathcal{R}_1(i) \cap \mathcal{R}_1(i') = \{j\}$ . Then

$$M_{ij} = M_{i'j} = P_{ij} + P_{i'j} = \frac{1}{|f'(x_{ij})|} \frac{m(X_j \cap fX_i)}{m(X_j)} + \frac{1}{|f'(x_{i'j})|} \frac{m(X_j \cap fX_{i'})}{m(X_j)}.$$

Since  $X_j \subset f(X_i) \cup f(X_{i'})$ , this implies that  $M_{ij}$  can be written as a convex combination of  $\frac{1}{|f'(x_{ij})|}$  and  $\frac{1}{|f'(x_{i'j})|}$ . By the intermediate value property for  $f'$ , there exists another point  $x$  in the interval with endpoints  $x_{ij}$  and  $x_{i'j}$  such that  $M_{ij} = 1/|f'(x)|$ . By expansivity,

$\rho(x_{ij}, X_{i'})$ ,  $\rho(x_{i'j}, X_i) \leq \frac{2^{-N}}{\lambda}$  and hence  $\rho(x, X_i)$ ,  $\rho(x, X_{i'}) \leq \frac{2^{-N}}{\lambda}$ , completing the proof of the claim.

To see why (8) follows from the claim, observe that if  $z_1 \in X_{i_1}$  and  $z_2 \in X_{i_2}$ , then  $\rho(z_1, z_2) \leq 2^{-N}(|i_1 - i_2| + 1)$ . Now suppose that  $M_{i_1 j_1}, M_{i_2 j_2} > 0$ . Then there exist  $x_1$  and  $x_2$  such that  $M_{i_1 j_1} = \frac{1}{|f'(x_1)|}$  and  $M_{i_2 j_2} = \frac{1}{|f'(x_2)|}$ , and both  $\rho(x_1, X_{i_1}), \rho(x_2, X_{i_2}) \leq \frac{2^{-N}}{\lambda}$ . Hence  $\rho(x_1, x_2) \leq 2^{-N}(|i_1 - i_2| + 1 + 2/\lambda)$ , and (8) follows from the distortion inequality (2).  $\square$

**Lemma 3.1.** *For every  $i \in \mathbb{Z}$  and  $n > 0$ , we have that*

$$P_n(k; i) \leq \lambda^{-n} \quad \forall k \in \mathcal{R}_n(i).$$

**Proof.** This is easily proved by an induction exploiting (7) and the construction of  $M$ ; cf the proof of lemma 3.3 below.  $\square$

Having collected the necessary facts about the Markov chain, we recall some useful arithmetic facts.

**Lemma 3.2 (simple facts).** Let  $\{x_i\}_{i=1}^n$  and  $\{y_i\}_{i=1}^n$  be two collections of positive real numbers. Then:

(1) if  $c_1 = \min_{i=1}^n \frac{x_i}{y_i}$  and  $c_2 = \max_{i=1}^n \frac{x_i}{y_i}$  then

$$c_1 \leq \frac{\sum_{i=1}^n x_i}{\sum_{i=1}^n y_i} \leq c_2$$

(2) if  $c_0 \geq 1$  is such that  $\frac{x_i}{x_j} \leq c_0$  for all  $i, j = 1 \dots n$  then for each  $k$

$$\frac{\sum_{i=1}^n x_i}{c_0 n} \leq x_k \leq c_0 \frac{\sum_{i=1}^n x_i}{n}.$$

**Proof.** Easy.  $\square$

### 3.2. Rows of $P^N$

Our aim in this section is to prove a uniform rate of exponential convergence to equilibrium for rows of powers of the transition matrix  $P^N$ . This result is theorem 2, and below we prove the main element (proposition 2).

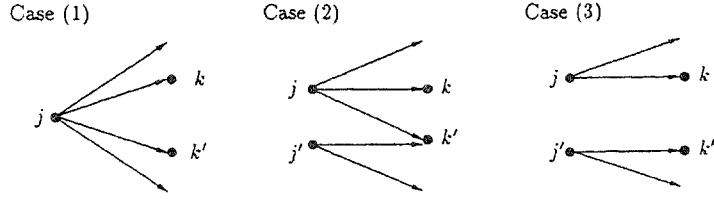
We begin by passing to the lift of the Markov chain. For a fixed initial state  $i_0$ , we estimate the distortion of the distribution  $P_N(\cdot; i_0)$ , and use this to obtain explicit bounds on the distribution when it is projected back down onto  $\mathbb{Z}_{2^N}$ .

First, we control the distribution restricted to  $\mathcal{R}_n^*(i)$  for an arbitrary natural number  $n$ , and state  $i \in \mathbb{Z}$ . We then show that this implies a uniform lower bound on the rows of  $P^N$ ; from this we can deduce an explicit bound on the exponential rate of decay of correlations for  $P^N$ .

**Lemma 3.3.** *Let  $N \geq N_0$ ,  $n > 0$  and suppose that  $k, k' \in \mathcal{R}_n^*(i)$  are adjacent. Then*

$$\frac{P_n(k; i)}{P_n(k'; i)} \leq e^{K\kappa n 2^{-N}}.$$





**Figure 1.** Possible transitions to adjacent states  $k$  and  $k'$  (arrows represent a positive transition probability).

**Proof.** Because  $N \geq N_0$ , (H1)–(H3) are satisfied. For  $n = 1$ , the lemma is given by (H3). We use induction.

Suppose that for all  $j, j' \in \mathcal{R}_m^*(i)$  satisfying  $|j - j'| = 1$

$$\frac{P_m(j; i)}{P_m(j'; i)} \leq e^{K\kappa m 2^{-N}}.$$

Let  $k, k' \in \mathcal{R}_{m+1}^*(i)$  be adjacent. Then we have the following possibilities, which are shown in figure 1.

These cases are exhaustive because of the continuity conditions (H1) and (H2). We give a proof only for case (2), because case (1) is similar (but easier) and case (3) is a limiting case of case (2).

Now, it is easy to see from figure 1 that

$$\begin{aligned} P_{m+1}(k; i) &= P_{jk} P_m(j; i) \\ P_{m+1}(k'; i) &= P_{jk'} P_m(j; i) + P_{j'k'} P_m(j'; i). \end{aligned} \quad (10)$$

It follows from (H1) and (H2) that

$$k \in \mathcal{R}_1^*(j) \quad \text{and} \quad k' \in \mathcal{R}_1(j) \cap \mathcal{R}_1(j')$$

and consequently that

$$M_{jk} = P_{jk} \quad \text{and} \quad M_{j'k'} = P_{jk} + P_{j'k'}. \quad (11)$$

Since  $j, j' \in \mathcal{R}_m^*(i)$ , the induction hypothesis implies that

$$P_m(j'; i) \leq P_m(j; i) e^{K\kappa m 2^{-N}}. \quad (12)$$

Equations (10) and (12) together imply

$$P_{m+1}(k'; i) \leq P_m(j; i) e^{K\kappa m 2^{-N}} M_{j'k'} \quad (13)$$

and hence

$$\frac{P_{m+1}(k'; i)}{P_{m+1}(k; i)} \leq \frac{P_m(j; i) e^{K\kappa m 2^{-N}} M_{j'k'}}{P_m(j; i) M_{jk}} \leq e^{K\kappa(m+1)2^{-N}}.$$

The last inequality follows from equation (8) in (H3), and the lemma follows by induction on  $m$ .  $\square$

Proposition 1 informs us about the *regularity* of the distribution  $P_n(\cdot; i)|_{\mathcal{R}_n^*(i)}$ .

**Proposition 1.** Let  $N \geq N_0$ . Suppose that  $n \leq N$ , and that the  $k, k' \in \mathcal{R}_n^*(i)$  for some  $i \in \mathbb{Z}$  are such that  $|k - k'| < 2^N$ . Define

$$c_n = e^{K\kappa(n^2 + 3n + 1)2^{-N} + K/(\lambda - 1)}$$

then,

$$\frac{P_n(k; i)}{P_n(k'; i)} \leq c_n.$$

In particular, if  $C_0 = e^{K\kappa(N_0^2+3N_0+1)2^{-N_0+K/(\lambda-1)}}$  then for every  $n \leq N$  we have independently of  $N$  that

$$\frac{P_n(k; i)}{P_n(k'; i)} \leq C_0$$

for all  $k, k' \in \mathcal{R}_n^*(i)$  with  $|k - k'| < 2^N$ .

**Proof.** By an argument similar to the proof of lemma 3.3, we have that given  $k, k' \in \mathcal{R}_n^*(i)$ , there exist  $j, j' \in \mathcal{R}_{n-1}^*(i)$  satisfying

$$\begin{aligned} P_{jk} > 0 & \quad P_n(k; i) \leq P_{n-1}(j; i)M_{jk}e^{K\kappa n2^{-N}} \\ P_{j'k'} > 0 & \quad P_n(k'; i) \geq P_{n-1}(j'; i)M_{j'k'}e^{-K\kappa n2^{-N}} \end{aligned}$$

(compare with equation (13)). Moreover, we have from (H3) that

$$|j - j'| \leq \frac{|k - k'| + 1}{\lambda} + 1$$

and consequently (again from (H3))

$$\frac{M_{jk}}{M_{j'k'}} \leq e^{K2^{-N}(\frac{|k-k'|+1}{\lambda}+1+\kappa)}.$$

Putting all this together:

$$\frac{P_n(k; i)}{P_n(k'; i)} \leq e^{K2^{-N}(\kappa(2n+1)+\frac{|k-k'|+1}{\lambda}+1)} \frac{P_{n-1}(k_{n-1}; i)}{P_{n-1}(k'_{n-1}; i)}$$

(where we have relabelled  $j = k_{n-1}$  and  $j' = k'_{n-1}$ ). We now repeat this argument inductively to obtain (for  $m \leq n$ )

$$\frac{P_n(k; i)}{P_n(k'; i)} \leq \frac{P_{n-m}(k_{n-m}; i)}{P_{n-m}(k'_{n-m}; i)} e^{K2^{-N}(\kappa \sum_{l=n-m+1}^n (2l+1) + \sum_{l=1}^m \frac{|k-k'|}{\lambda^l} + \sum_{l=0}^m \frac{m+1-l}{\lambda^l} - 1)} \quad (14)$$

with  $k_{n-m}, k'_{n-m} \in \mathcal{R}_{n-m}^*(i)$  satisfying  $|k_{n-m} - k'_{n-m}| \leq \frac{|k-k'|}{\lambda^m} + \sum_{l=0}^m \lambda^{-l}$ . Because  $\mathcal{R}_0^*(i) = \{i\}$ , it is clear (in the notation of (14)) that  $k_0 = k'_0 = i$ . Finally, since we assumed that  $|k - k'| < 2^N$ , we can rewrite (14) for  $m = n$ :

$$\frac{P_n(k; i)}{P_n(k'; i)} \leq e^{K2^{-N}(\kappa \sum_{l=1}^n (2l+1) + 2^N \sum_{l=1}^n \lambda^{-l} + \sum_{l=0}^n \frac{n+1-l}{\lambda^l} - 1)}.$$

It is easy to check that the exponential is bounded by the constant  $c_n$  defined in the statement of the theorem.

The last part follows because  $c_n \leq c_N$  whenever  $n \leq N$ , and  $c_N \leq C_0$  (because  $N_0 > 1$ ).  $\square$

**Remark 3.1.** Observe that  $\max_{n \leq N} \{c_n\} \rightarrow e^{K/(\lambda-1)}$  as  $N \rightarrow \infty$ . Therefore if we put  $C(N) = \max_{n \leq N} \{c_n\}$ , then for fixed  $N$  the conclusion of the proposition may be replaced by

$$\frac{P_n(k; i)}{P_n(k'; i)} \leq C(N)$$

for all  $i \in \mathbb{Z}_{2^n}$ ,  $k, k' \in \mathcal{R}_n^*(i)$  with  $|k - k'| < 2^n$ ,  $n \leq N$ . Hence, for large  $N$ , proposition 1 implies a distortion inequality for the contribution of  $\mathcal{R}_n^*(i)$  to the rows of  $P^n$  which is comparable with the regularity of the invariant density; cf the folklore theorem. Indeed one can deduce from equation (14) in the proof that *asymptotically*  $P_n(k; i)/P_n(k'; i) \leq e^{K2^{-N}|k-k'|/(\lambda-1)} + \epsilon_N$ , where the correction  $\epsilon_N$  is exponentially small in  $N$ .

Proposition 1 gives control over the distribution  $P_n(\cdot; i)$  restricted to  $\mathcal{R}_n^*(i)$ . However, because all of  $\mathcal{R}_n(i)$  contributes to the  $i$ th row of  $P^n$ , we now show that  $\mathcal{R}_n(i)$  can be decomposed as a collection of sets like  $\mathcal{R}_n^*(i)$ .

**Lemma 3.4 (decomposition of  $\mathcal{R}_n(i)$ ).** *Fix  $n$ . For each  $j = 0, \dots, n$ , there exists a finite collection of states  $\{i_r^{(j)}\}_{r=1}^{2^j} \subset \mathbb{Z}$  such that:*

- (1)  $r_j \leq 2^j$ ;
- (2) each  $i_r^{(j)} \in \mathcal{R}_j(i)$  (we make the obvious labelling  $i_r^{(0)} = i$ );
- (3)

$$\mathcal{R}_n(i) = \bigcup_{j=0}^n \bigcup_{r=1}^{2^j} \mathcal{R}_{n-j}^*(i_r^{(j)}).$$

We call each  $\mathcal{R}_{n-j}^*(i_r^{(j)})$  a component of the decomposition.

**Proof.** The lemma follows more or less directly from (H1) and (H2). To help understand the decomposition, we elaborate the construction. First, we define *reachability* from an interval of states in  $\mathbb{Z}$ . Let  $J \subset \mathbb{Z}$  be an interval of states and denote

$$\mathcal{R}_m(J) = \bigcup_{j \in J} \mathcal{R}_m(j)$$

and

$$\mathcal{R}_m^*(J) = \mathcal{R}_m(J) \setminus \mathcal{R}_m(\mathbb{Z} \setminus J)$$

in complete analogy with definitions in equations (4) and (5). The comments in (H2) about the endpoints of images of adjacent states apply identically to the images of adjacent intervals. That is, if  $J_1$  and  $J_2$  are two adjacent intervals in  $\mathbb{Z}$  (by this we mean that there are no integers between the greatest element of  $J_1$  and the least element of  $J_2$ ) then either  $\exists j \in \mathbb{Z}$  such that

$$\mathcal{R}_1(J_1) \cap \mathcal{R}_1(J_2) = \{j\}$$

or the intersection is empty. Hence, if  $J_1, \dots, J_k$  is a family of adjacent intervals, then

$$\mathcal{R}_1(J_1 \cup \dots \cup J_k) = \mathcal{R}_1^*(J_1) \cup \dots \cup \mathcal{R}_1^*(J_k) \cup \{i_r\}_{r=0}^k$$

where each  $i_r \in \mathcal{R}_1(J_r) \cap \mathcal{R}_1(J_{r+1})$  (with the obvious interpretation in the cases where  $r = 0$  or  $k$ ). The final preparatory observation is that  $\mathcal{R}_1^*(\mathcal{R}_m^*(J)) = \mathcal{R}_{m+1}^*(J)$  for any interval of states  $J$  (this follows from the continuity hypotheses (H1) and (H2)).

Now, we construct the decomposition inductively. As in the comments following (H2), let  $\mathcal{R}_1(i) \setminus \mathcal{R}_1^*(i) = \{i_1^{(1)}, i_2^{(1)}\}$ , and write

$$\mathcal{R}_1(i) = \mathcal{R}_0^*(i_1^{(1)}) \cup \mathcal{R}_1^*(i) \cup \mathcal{R}_0^*(i_2^{(1)}).$$

This provides a basis for our induction.

Suppose the conclusion of the lemma is satisfied for  $j = m$ . Then the decomposition contains at most  $2^{m+1} - 1$  components (each component is obviously an interval):

$$\mathcal{R}_m(i) = \bigcup_{j=0}^m \bigcup_{r=1}^{2^j} \mathcal{R}_{m-j}^*(i_r^{(j)}).$$

By our previous remarks, there exist a sequence of states  $\{i_r^{(m+1)}\}$  containing at most  $2^{m+1}$  elements such that

$$\mathcal{R}_1(\mathcal{R}_m(i)) = \left\{ \bigcup_{j=0}^m \bigcup_{r=1}^{2^j} \mathcal{R}_1^*(\mathcal{R}_{m-j}^*(i_r^{(j)})) \right\} \cup \left\{ \bigcup_{r=1}^{2^{m+1}} \mathcal{R}_0^*(i_r^{(m+1)}) \right\}.$$

This equation can be rewritten to be exactly the conclusion of the lemma for  $j = m + 1$ , so the result follows by induction.  $\square$

In view of lemma 3.4, proposition 1 implies that  $P_n(\cdot; i)$  is well behaved on all the components  $\mathcal{R}_{n-j}^*(i_r^{(j)})$ . Now, we pick an initial state  $i_0$ , and fix  $n = N$ . To prove uniform bounds on the  $i_0$ th row of  $P^N$ , we pick out a subset  $\Lambda \subset \mathcal{R}_N(i_0)$  which contains ‘most’ of the distribution  $P_N(\cdot; i_0)$ , and is particularly amenable to the application of proposition 1 and lemma 3.4.

**Lemma 3.5.** *There exist disjoint subsets  $L_1, L_2 \subset \mathbb{Z}$  (depending on  $i_0$ ) with the following properties:*

- (1) if  $k \in \mathcal{R}_N(i_0)$ , then  $\exists l \in L_1 \cup L_2, \hat{k} \in \mathbb{Z}_{2^N}$  with  $k = \hat{k} + l2^N$ ;
- (2) if  $l \in L_1$  then there exists a unique  $i = i(l)$  such that

$$i(l) \in \bigcup_{j=0}^N \bigcup_{r=1}^{r_j} \{i_r^{(j)}\} \quad \text{and} \quad \hat{k} + l2^N \in \mathcal{R}_{N-j}^*(i(l)) \forall \hat{k} \in \mathbb{Z}_{2^N}$$

- (3)  $\text{card}\{L_2\} \leq 2^{N+1}$ ;
- (4)  $P_N(\Lambda; i_0) \geq (1 - 2(\frac{4}{\lambda})^N)$ , where  $\Lambda = \bigcup_{l \in L_1} \bigcup_{\hat{k} \in \mathbb{Z}_{2^N}} \{\hat{k} + l2^N\}$ .

**Proof.** Put

$$L_1 = \{l \in \mathbb{Z} : \exists j \leq N, i(l) \in \mathcal{R}_j(i_0) \text{ s.t. } \hat{k} + l2^N \in \mathcal{R}_{N-j}^*(i(l)) \forall \hat{k} \in \mathbb{Z}_{2^N}\}$$

and

$$L_2 = \{l \in \mathbb{Z} : \exists \hat{k} \in \mathbb{Z}_{2^N} \text{ s.t. } \hat{k} + l2^N \in \mathcal{R}_N(i_0)\} \setminus L_1.$$

The set  $L_1$  indexes those lifted copies of  $\mathbb{Z}_{2^N}$  which are wholly contained in one of the components  $\mathcal{R}_{N-j}^*(i_r^{(j)})$  of the decomposition in lemma 3.4. Properties (1) and (2) are obvious from the definition.

To see why (3) is true, notice that  $L_2$  consists of those  $l$  such that the ‘copy’  $\mathbb{Z}_{2^N} + l2^N$  of  $\mathbb{Z}_{2^N}$  overlaps more than one interval in the decomposition in lemma 3.4. Because  $\mathcal{R}_N(i_0)$  can be written as both a sequence of adjacent intervals from the above decomposition, and a sequence of adjacent copies of  $\mathbb{Z}_{2^N}$ , the number of copies of  $\mathbb{Z}_{2^N}$  which overlap more than one interval must be bounded by the number of boundary points of intervals from the decomposition. Lemma 3.4 now implies (3).

Finally,

$$P_N(\Lambda; i_0) = 1 - \sum_{l \in L_2} \sum_{\hat{k} \in \mathbb{Z}_{2^N}} P_N(\hat{k} + l2^N; i_0).$$

Since each  $P_N(k; i_0) \leq \lambda^{-N}$  (lemma 3.4), the double sum is bounded above by  $2^{N+1} \times 2^N \times \lambda^{-N}$ ; (4) follows.  $\square$

Recall from remark 1.2 that  $N_0 = \lceil \max\{\log \lambda_{\max}/\log 2, \log 3/\log \frac{\lambda}{4}\} \rceil$ , so that whenever  $N \geq N_0$ , we have  $P_N(\Lambda; i_0) \geq 1 - 2(\frac{4}{\lambda})^N \geq \frac{1}{3}$ . By construction, proposition 1 applies to all of  $\Lambda$ .

**Proposition 2 (uniform lower bound for  $P^N$ ).** *Let  $N \geq N_0$ ,  $P$  be the  $2^N \times 2^N$  area overlaps matrix, and  $C_0$  the constant (independent of  $N$ ) from proposition 1. For any two states  $i, k \in \mathbb{Z}_{2^N}$ ,*

$$\delta 2^{-N} \leq P_{ik}^N$$

where  $\delta = (1 - 2(\frac{4}{\lambda})^N)/C_0 \geq 1/(3C_0)$ .

**Proof.** Let  $i_0 = i$ , and  $L_1, L_2, \Lambda$  be as in lemma 3.5. Let  $k' \in \mathbb{Z}_{2^N}$  (so that  $|k - k'| \leq 2^N$ ), and recall from equation (6) that

$$P_{i_0 k}^N = \sum_{l \in L_1 \cup L_2} P_N(k + l 2^N; i_0) \geq \sum_{l \in L_1} P_N(k + l 2^N; i_0)$$

(a similar formula holds for  $k'$ ). Next, we observe that proposition 1 allows us to control the distortion of the distribution  $P_N(\cdot; i_0)$  within components such as those given in the decomposition in lemma 3.4. More precisely, for each  $l \in L_1$  let  $j \leq N$ , and  $i(l) \in \mathcal{R}_j(i_0)$  be such that  $k + l 2^N, k' + l 2^N \in \mathcal{R}_{N-j}^*(i(l))$ . Then

$$\begin{aligned} P_N(k + l 2^N; i_0) &= P_j(i(l); i_0) P_{N-j}(k + l 2^N; i_r^{(j)}) \\ P_N(k' + l 2^N; i_0) &= P_j(i(l); i_0) P_{N-j}(k' + l 2^N; i_r^{(j)}). \end{aligned}$$

Because  $|k + l 2^N - (k' + l 2^N)| = |k - k'| \leq 2^N$ , proposition 1 (with  $n = N - j$  and  $i = i(l)$ ) implies

$$\frac{P_N(k + l 2^N; i_0)}{P_N(k' + l 2^N; i_0)} \leq C_0. \quad (15)$$

Since equation (15) applies to each  $l \in L_1$ , we have from lemma 3.2:

$$\frac{\sum_{l \in L_1} P_N(k + l 2^N; i_0)}{\sum_{l \in L_1} P_N(k' + l 2^N; i_0)} \leq \max_{l \in L_1} \frac{P_N(k + l 2^N; i_0)}{P_N(k' + l 2^N; i_0)} \leq C_0.$$

Since  $P_N(\Lambda; i_0) = \sum_{k' \in \mathbb{Z}_{2^N}} \sum_{l \in L_1} P_N(k' + l 2^N; i_0)$ , lemmas 3.2 and 3.5 imply

$$\sum_{l \in L_1} P_N(k + l 2^N; i_0) \geq \frac{1}{2^N C_0} P_N(\Lambda; i_0) \geq \frac{1}{2^N C_0} \left(1 - 2\left(\frac{4}{\lambda}\right)^N\right). \quad (16)$$

The result follows.  $\square$

**Remark 3.2.** We emphasize that the constant  $\delta$  is such that the conclusion of proposition 2 is *uniform* in  $i, k$ , and  $N \geq N_0$ .

In fact, for large  $N$  we can get an even better estimate of  $\delta$  by applying remark 3.1. Rather than using the constant  $C_0$  (which works for all  $N \geq N_0$ ), let  $C(N)$  be as in that remark. Then proposition 2 is true with  $C(N)$  replacing  $C_0$  in the definition of  $\delta$ . Since remark 3.1 shows that  $C(N) \rightarrow e^{k/(\lambda-1)}$  exponentially fast in  $N$ , we can replace  $1/(3C_0)$  with

$$\delta' = e^{-K/(\lambda-1)} + \epsilon_N$$

in the conclusion of the proposition (the correction  $\epsilon_N$  is exponentially small in  $N$ ; cf. remark 3.1 and equation (1)).

Although it plays no further part in the argument, we include for completeness a uniform distortion property for rows of  $P^N$ .

**Corollary 3.6 (distortion inequality for rows of  $P^N$ ).** *Let  $C_0$  be as in proposition 1. Let  $i, k, k' \in \mathbb{Z}_{2^N}$  be any states in the Markov chain with transition matrix  $P$ . Then*

$$\frac{P_{ik}^N}{P_{ik'}^N} \leq \frac{C_0}{1 - 2(\frac{4}{\lambda})^N}.$$

**Proof.** Let  $L_1$  and  $L_2$  be the sets in lemma 3.5 and recall from lemmas 3.1 and 3.5 (3) that

$$P_N(k + l2^N; i_0) \leq \lambda^{-N} \forall l \in L_2 \quad \text{and} \quad \text{card}\{L_2\} \leq 2^{N+1}.$$

Thus,

$$\begin{aligned} \frac{P_{ik}^N}{P_{ik'}^N} &= \frac{\sum_{l \in (L_1 \cup L_2)} P_N(k + l2^N; i)}{\sum_{l \in (L_1 \cup L_2)} P_N(k' + l2^N; i)} \\ &\leq \frac{\sum_{l \in L_1} P_N(k + l2^N; i)}{\sum_{l \in L_1} P_N(k' + l2^N; i)} + \frac{\sum_{l \in L_2} P_N(k + l2^N; i)}{\sum_{l \in L_1} P_N(k' + l2^N; i)} \\ &\leq \max_{l \in L_1} \frac{P_N(k + l2^N; i)}{P_N(k' + l2^N; i)} + \frac{2^{N+1} \times \lambda^{-N}}{\sum_{l \in L_1} P_N(k' + l2^N; i)}. \end{aligned}$$

Then, equations (15) and (16) from the proof of proposition 2 imply that

$$\frac{P_{ik}^N}{P_{ik'}^N} \leq C_0 + \frac{2^{N+1} \lambda^{-N}}{2^{-N}(1 - 2(\frac{4}{\lambda})^N)/C_0} = \frac{C_0}{1 - 2(\frac{4}{\lambda})^N}.$$

□

**Remark 3.3.** In fact, putting together remarks 3.1 and 3.2 we have that

$$\frac{P_{ik}^N}{P_{ik'}^N} \leq e^{K2^{-N}|k-k'|/(\lambda-1)} + \epsilon_N$$

where the error  $\epsilon_N$  is exponentially small in  $N$ .

### 3.3. Convergence of rows of $P^{kN}$ to equilibrium

In this brief section, we show how to use proposition 2 to prove an exponential decay of correlations property for  $P$ ; this is theorem 2.

**Lemma 3.7.** *Suppose that  $P$  is a stochastic matrix, that  $r = (r_j)$  is a probability vector, and that  $\alpha \in (0, 1)$  is such that*

$$\alpha r_j \leq P_{ij}^N$$

for each  $i, j$ . Then for any pair  $x, y$  of probability vectors,

$$\|xP^{kN+l} - yP^{kN+l}\|_1 \leq \|x - y\|_1 (1 - \alpha)^k$$

for  $k, l \geq 0$ , and  $\|\cdot\|_1$  the usual vector 1-norm.

**Remark 3.4.** This lemma is a version of Doeblin's condition and is an easy corollary to theorem 16.2.4 in [7].

**Theorem 2 (speed of convergence to equilibrium).** *Let  $N \geq N_0$ . Let  $P$  be the  $2^N \times 2^N$  area overlaps matrix. Then there exists  $\delta \in (0, 1)$ , independent of  $N$ , such that for any two probability vectors  $x, y \in \mathbb{R}_{\text{prob}}^{2^N}$ ,  $k > 0$ ,  $0 \leq l < N$*

$$\|(x - y)P^{kN+l}\|_1 \leq \|x - y\|_1(1 - \delta)^k.$$

*The constant  $\delta$  is given in proposition 2.*

**Proof.** The theorem follows directly from lemma 3.7 by using  $\alpha \equiv \delta$  (from proposition 2) and  $(r_i)_{i=1}^{2^N} \equiv (2^{-N}, \dots, 2^{-N})$ .  $\square$

**Remark 3.5.** We reiterate that the decay rate  $(1 - \delta)$  can be chosen independently of  $N$  (the logarithm of the size of the matrix); this is guaranteed by our choice of constants  $C_0$  and  $N_0$ . In fact, by combining the conclusions of propositions 1 and 2 we can write down an explicit formula for  $\delta$ . Rather than do this here, we note that for large  $N$ , the conclusion of theorem 2 is valid for  $\delta$  exponentially close to  $e^{-K/(\lambda-1)}$  (remark 3.2).

#### 4. Proof of theorem 1

##### 4.1. The invariant density and another Markov approximation

The area overlaps Markov chain is constructed from the Lebesgue measure  $m$  on the circle  $S^1$ . To analyse the proximity of the invariant vector for this Markov chain to the acim, we compare the area overlaps Markov chain with a similar Markov approximation based on the acim  $\mu$ . First, we show that  $\mu$  is locally ‘like’ the Lebesgue measure.

**Lemma 4.1.** *Let  $f, \mu$  be as described, and let  $A$  be any interval in  $S^1$ . If  $B \subset A$ , then*

$$\frac{1}{C_A} \frac{m(B)}{m(A)} \leq \frac{\mu(B)}{\mu(A)} \leq C_A \frac{m(B)}{m(A)}$$

where  $C_A = e^{\frac{K \text{diam}(A)}{\lambda-1}}$  ( $\text{diam}(A)$  is the length of the interval  $A$ ).

**Proof.** From the folklore theorem, if  $x, y \in A$  then

$$\frac{\phi(x)}{\phi(y)} \leq C_A.$$

We use this distortion inequality to compare the density  $\phi$  with the Lebesgue measure. Pick  $x_0 \in \bar{A}$  such that

$$\phi(x_0) = \sup_{x \in A} \phi(x).$$

We obviously have for each point  $x \in A$ ,

$$\frac{\phi(x_0)}{C_A} \leq \phi(x) \leq \phi(x_0).$$

Consequently, by integrating over an arbitrary subset  $B \subset A$

$$\frac{\phi(x_0)}{C_A} m(B) \leq \mu(B) \leq \phi(x_0) m(B)$$

and

$$\frac{\phi(x_0)}{C_A} m(A) \leq \mu(A) \leq \phi(x_0) m(A).$$

The result follows immediately from these inequalities.  $\square$

Next, we define another Markov chain. Let a transition matrix  $Q$  be defined by putting

$$Q_{ij} = \frac{\mu(f^{-1}X_j \cap X_i)}{\mu(X_i)}.$$

It follows easily that the vector  $q = (q_i) = (\mu(X_i))_{i=1}^n$  is an invariant vector for the corresponding Markov chain. The results in the previous section tell us about the rate of convergence to equilibrium of the Markov chain governed by  $P$ . In the next section we apply this to show that ‘nearby’ Markov chains have ‘nearby’ invariant vectors. The remaining two results in this section show that  $P$  and  $Q$  are close.

**Corollary 4.2.** *Let  $f$ ,  $C_A$  be as in lemma 4.1, except that  $A = X_i$  is an element of the partition over which  $P$  and  $Q$  are constructed. Then*

$$\frac{1}{C_N} P_{ij} \leq Q_{ij} \leq C_N P_{ij}$$

where  $C_N = e^{\frac{K2^{-N}}{\lambda-1}}$ .

**Proof.** Apply lemma 4.1 with  $B = f^{-1}X_j \cap X_i$ .  $\square$

Hence, we have the following.

**Lemma 4.3.** *For any probability vector  $x$ , we have*

$$\|x(P - Q)\|_1 \leq C_N - 1 \approx \frac{K2^{-N}}{\lambda - 1}.$$

**Proof.** By applying corollary 4.2 we obtain

$$\frac{1}{C_N} (xP)_j = \frac{1}{C_N} \sum_{i=1}^n x_i P_{ij} \leq \sum_{i=1}^n x_i Q_{ij} = (xQ)_j \leq C_N (xP)_j.$$

Then, since  $C_N > 1$ ,

$$|(xQ)_j - (xP)_j| \leq (C_N - 1) (xP)_j$$

and since  $xP$  is also a probability vector, summing over  $j$  gives the result.  $\square$

#### 4.2. The main result

Now, we have an exponential convergence to equilibrium property for  $P$  (theorem 2) and that a characterization of the ACIM (the vector  $q$ ) arises as the invariant vector of a Markov chain with transition probabilities close to  $P$  (corollary 4.2).

**Theorem 3 (restatement of the main theorem).** *Let  $N_0$  be as defined in remark 1.2 and suppose that  $N \geq N_0$ . Let  $P$  be the  $2^N \times 2^N$  area overlaps matrix. Let  $\delta \in (0, 1)$  be the constant from theorem 2 and let  $p = (p_i)$  be the invariant probability vector for  $P$ . If  $\mu$  is the unique absolutely continuous invariant measure for  $f$ , and  $q$  is the probability vector with components  $q_i = \mu(X_i)$ , then*

$$\sum_{i=1}^{2^N} |\mu(X_i) - p_i| = \|q - p\|_1 \leq \frac{N}{\delta} (e^{\frac{K2^{-N}}{\lambda-1}} - 1) \approx C N 2^{-N}$$

where the approximate equality holds for large  $N$ , and  $C = \frac{1}{\delta} \frac{K}{\lambda-1}$ . We give a formula for  $\delta$  below.



**Proof.** First of all, it follows easily from lemma 3.7 that for any probability vector  $x$ ,

$$p = \lim_{n \rightarrow \infty} x P^n.$$

In particular, we can put  $x = q$ , to obtain

$$\|q - p\|_1 \leq \sum_{n=0}^{\infty} \|q P^n - q P^{n+1}\|_1.$$

But we can use the fact that  $qQ = q$  to rewrite this as

$$\|q - p\|_1 \leq \sum_{k=0}^{\infty} \sum_{l=0}^{N-1} \|(qQ - qP) P^{kN+l}\|_1$$

which, in view of theorem 2, is bounded above by

$$\sum_{k=0}^{\infty} N(1 - \delta)^k \|qQ - qP\|_1.$$

The theorem now follows from lemma 4.3. □

**Remark 4.1.** The number  $N_0 = \lceil \max\{\log_2 \lambda_{\max}, \log 3 / \log \frac{\lambda}{4}\} \rceil$  was chosen to guarantee that the lift of the Markov chain is well defined (cf section 2), and to ensure that  $\delta$  in the theorem is bounded below. From propositions 1 and 2 we obtain

$$\delta \geq \left(1 - 2 \left(\frac{4}{\lambda}\right)^{N_0}\right) e^{-K \frac{\lambda+2}{\lambda} (N_0^2 + 3N_0 + 1) 2^{-N_0}} e^{-K/(\lambda-1)} \geq \frac{1}{3} e^{-K \frac{\lambda+2}{\lambda} (N_0^2 + 3N_0 + 1) 2^{-N_0}} e^{-K/(\lambda-1)}.$$

(The constants  $K$  and  $\lambda$  are as in equations (1) and (2).)

In fact, by choosing a value of  $N_0$  which is *larger* than the minimal  $N_0$  defined above, we can get much smaller estimates of the constants in the theorem: For each  $N_1 \geq N_0$  put

$$C_1(N_1) = e^{K \frac{\lambda+2}{\lambda} (N_1^2 + 3N_1 + 1) 2^{-N_1} + K/(\lambda-1)} \quad \text{and} \quad \beta(N_1) = 1 - 2(4/\lambda)^{N_1}.$$

Because  $\lambda > 4$ , we have that  $C_1(N_1)$  *decreases* as a function of  $N_1$ , and that  $\beta(N_1)$  *increases* as a function of  $N_1$  (the choice of  $N_0$  is partially explained by the fact that we always have  $\beta(N_0) \geq \frac{1}{3}$ .) Finally, we put

$$C_2(N_1) = \frac{C_1(N_1)}{\beta(N_1)} 2^{N_1} (e^{K 2^{-N_1}/(\lambda-1)} - 1).$$

Replacing  $N_0$  with  $N_1$  in the hypothesis of the theorem, we can replace  $C$  by  $C_2(N_1)$  in the conclusion.

Finally, note that  $C_2(N_1)$  *decreases* as  $N_1$  increases. In fact,

$$C_2(N_1) = \frac{K}{\lambda - 1} e^{K/(\lambda-1)} + \epsilon(N_1)$$

where the correction term  $\epsilon(N_1)$  is exponentially small in  $N_1$ . In practice, this limiting value is approached very quickly.

**Table 1.** Expansivity constants and acim approximation error bounds for certain maps  $g_l$ . The expansivity and distortion constants are computed by a fine numerical search (so are really just accurate estimates), and the constant  $C(l, n)$  is calculated as in [6]. The final column gives a bound on  $\|p_l^{(n)} - q_l^{(n)}\|_1$ , where  $p_l^{(n)}$  is the computed approximation and  $q_l^{(n)}$  is the vector of the exact invariant measure for  $g_l$ .

$l$	$\lambda_l$	$K_l$	$e^{K_l 2^{-n}/(\lambda_l - 1)} - 1$	$C(l, n)$	$C(l, n) \times (e^{K_l 2^{-n}/(\lambda_l - 1)} - 1)$
2	5.66	14.8	$1.55 \times 10^{-3}$	10.5	0.0163
3	6.00	5.52	$5.39 \times 10^{-4}$	11.1	0.00599
4	4.75	6.10	$7.95 \times 10^{-4}$	12.2	0.00969
5	6.30	28.5	$2.62 \times 10^{-3}$	11.3	0.0297

## 5. Numerical example

We consider a family of unimodal expanding maps of the interval  $[0, 1]$ . For each  $l \geq 2$  let the maps  $f_l$  and  $h_l$  be given by

$$f_l(x) = 1 - |2x - 1|^l \quad \text{and} \quad h_l(x) = \frac{1}{2}(x^{1/l} + (1-x)^{1/l}).$$

The maps  $f_l$  are of theoretical interest, and have been well studied [1]. In particular, for each  $l$  there exists a unique absolutely continuous invariant measure  $\mu_l$ ; see [8]. The maps  $f_l$  are not uniformly expanding because each has a critical point at  $\frac{1}{2}$ , and except for the case of  $l = 2$  (the fully developed logistic map) no formula exists for the invariant density.

To illustrate our results, it can be shown that the *conjugate* family of maps  $\{g_l\}$  defined by  $g_l = h_l \circ f_l \circ h_l^{-1}$  are *uniformly expanding*. We approximate the acim for each  $f_l$  by applying the Markov scheme to  $g_l$ , and pulling back the approximate measure thus obtained by the conjugacy  $h_l$ . In fact, we apply the scheme to  $g_l^{m_l}$  where  $m_l \geq 1$  is chosen large enough that  $|(g_l^{m_l})'| > 4$ . For each  $l$ , we determine constants  $\lambda_l$  and  $K_l$  such that the uniform expansivity conditions (1) and (2) are satisfied for  $g_l^{m_l}$ .

Fix  $l$  (in the illustration below we have used  $l = 2, 3, 4, 5$ ). For each  $n = 2 \dots 11$ , the interval  $[0, 1]$  is divided into  $2^n$  equal subintervals, and the Markov approximation scheme is performed on  $g_l^{m_l}$ ; let the transition matrix be called  $P(l, n)$ . The approximate invariant measure for  $P(l, n)$  is denoted by  $p_l^{(n)}$ , and the restriction of the acim for  $g_l^{m_l}$  to the partition consisting of the  $2^n$  subintervals is denoted by  $q_l^{(n)}$ .

From the proof of theorem 3,

$$\|q_l^{(n)} - p_l^{(n)}\|_1 \leq C(l, n)(e^{K_l 2^{-n}/(\lambda_l - 1)} - 1) \quad (17)$$

where the constant  $C(l, n)$  satisfies

$$\sum_{k=0}^{\infty} \|(x - y)P^k\|_1 \leq \|x - y\|_1 C(l, n)$$

for any two probability vectors  $x, y$ . For each  $l, n$ , we numerically determine such constants<sup>†</sup>. These results are presented in table 1. Note that the right-hand column gives an *explicit numerical bound* on the accuracy of the measure approximation.

To explain the connection with the main result of this paper, observe that the proof of theorem 3 involves the observation that  $C(l, n)$  is bounded above by  $n/\delta_l$ , where  $\delta_l < 1$  is a constant depending only the map  $g_l^{m_l}$ ; cf proposition 2. In the remarks following theorem 3

<sup>†</sup> This is explained in more detail elsewhere [6]; but briefly, let  $\Gamma$  denote the set of all differences of probability vectors in  $\mathbb{R}^n$ , the idea of the calculation is to explicitly compute the sum for all the elements in a certain basis of  $\Gamma$ .

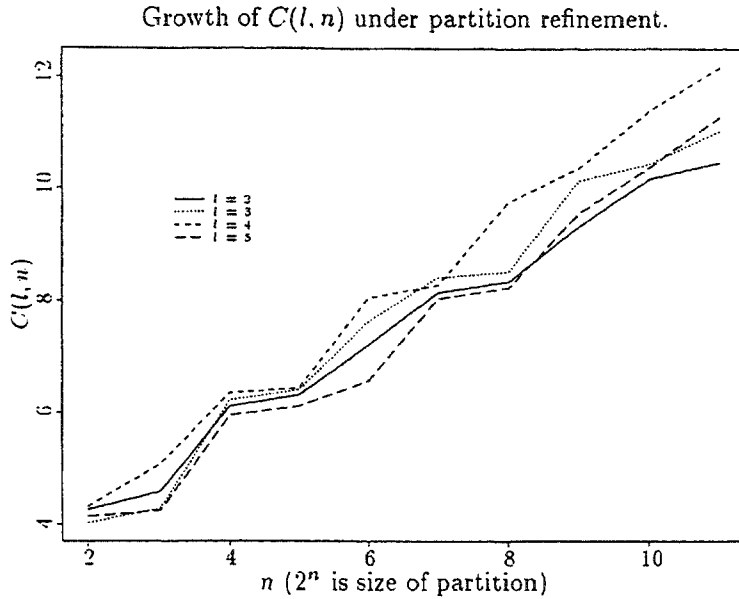


Figure 2. Scaling of error bound constants as partition refined.

Error bounds for measure approximations as partition refined.

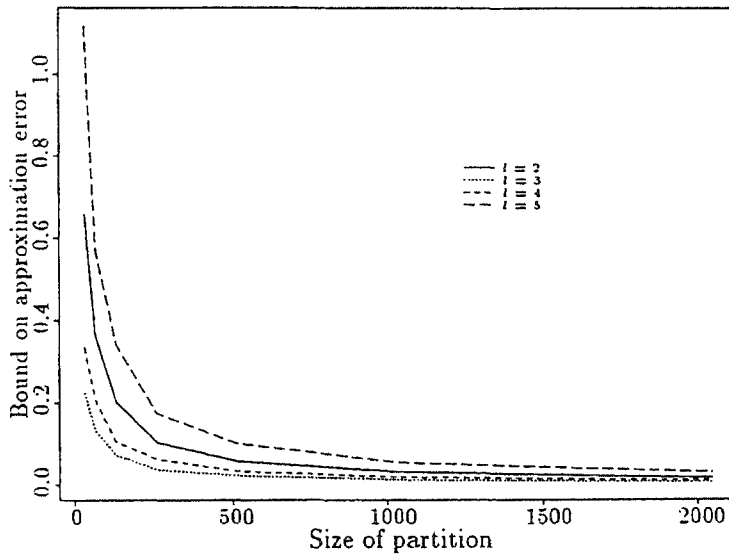


Figure 3. Decay of approximation error under partition refinement. For each  $l$ , the approximation error for the absolutely continuous invariant measure is expected to decay like  $O((\log m)/m)$  where  $m$  is the number of elements in the partition over which the approximation is constructed. This may be seen from the plot. Note that the bounds for  $m = 2048$  are precisely those appearing in the last column of table 1.

we observed that predicted bounds improve as  $n$  is increased. With this in mind, we present some numerical calculations for  $C(l, n)$ , rather than the theoretically determined bounds in

the theorem. For each  $l$ , we computed a bound on  $C(l, n)$  for  $n = 2 \dots 11$  (as in [6]). The predicted logarithmic growth in partition size (linear in  $n$ ) is borne out by the calculations displayed in figure 2.

Finally, we use the data for  $C(l, n)$  to illustrate that the behaviour predicted by theorem 3 can be observed numerically. Theorem 3 implies that  $\|q_l^{(n)} - p_l^{(n)}\|_1 \leq O(n2^{-n})$ . Having seen that  $C(l, n)$  grows logarithmically in partition size, we conclude the presentation of numerical results by plotting  $C(l, n)(e^{K_l 2^{-n}/(\lambda_l - 1)} - 1)$  (the upper bound on approximation error—cf equation (17)) against  $2^n$  (the partition size); this is figure 3.

### Acknowledgment

LSY was supported in part by the National Science Foundation.

### References

- [1] de Melo W and van Strien S 1993 *One-dimensional Dynamics* (Berlin: Springer)
- [2] Jiu Ding and Aihui Zhou 1996 Finite approximations of Frobenius–Perron operators. A solution of Ulam’s conjecture to multi-dimensional transformations *Physica D* **92** 61–8
- [3] Froyland G 1995 Finite approximation of Sinai–Bowen–Ruelle measures for Anosov systems in two dimensions *Random Comput Dynam.* **3** 251–64
- [4] Krzyżewski K and Szlenk W 1969 On invariant measures for expanding differentiable mappings *Stud. Math.* **33** 83–92
- [5] Tien-Yien Li 1976 Finite approximation for the Perron–Frobenius operator. A solution to Ulam’s conjecture *J. Approx. Theor.* **17** 177–86
- [6] Murray R D A 1996 Approximation and concrete error bounds for ACIM density calculation for  $C^{1+\alpha}$  expanding mappings *Statistical Laboratory Research Report* 96–22, University of Cambridge
- [7] Meyn S P and Tweedie R L 1993 *Markov Chains and Stochastic Stability* (Berlin: Springer)
- [8] Nowicki T and van Strien S 1991 Invariant measures exist under a summability condition for unimodal maps *Inv. Math.* **105** 123–36
- [9] Renyi A 1957 Representations of real numbers and their ergodic properties *Acta Math. Akad. Sc. Hungar.* **8** 477–93
- [10] Ulam S 1960 *Problems in Modern Mathematics* (New York: Interscience)
- [11] Young L-S 1995 Ergodic theory of differentiable dynamical systems *Real and Complex Dynamical Systems* ed B Branner and P Hjorth (Dordrecht: Kluwer) pp 293–336